IUT de Bordeaux
Département Informatique

M2201 – Graphes et Langages

Éléments de Théorie des Langages – Partie 1

- **Introduction générale : alphabets, mots et langages**
- **Langages rationnels**

Alphabet, mot sur un alphabet (1)

Un **alphabet** est un ensemble (fini) **A** de symboles appelés **lettres**.

Un **mot** **m** sur un alphabet **A** est une séquence (finie) de lettres prises dans **A** : $\mathbf{m = a_1 \ldots a_k}$. Ce mot est de **longueur** (nombre de lettres) k : $\mathbf{|m| = k}$.

Il existe un unique mot de longueur nulle, le **mot vide**, noté ε.

On note $\mathbf{A^*}$ l'ensemble de tous les mots construits sur l'alphabet **A**. On note $\mathbf{A^n}$ l'ensemble des mots de $\mathbf{A^*}$ de longueur **n**.

En particulier, $\mathbf{A^0 = \{ \varepsilon \}}$ (et donc $\mathbf{A^0 \neq \emptyset}$!) et $\mathbf{A^1 = A}$.

Alphabet, mot sur un alphabet (2)

Soit a une lettre de **A** et m un mot de **A***. Le **nombre d'occurrences** de la lettre a dans m, noté $|m|_a$, est le nombre de fois où la lettre a apparaît dans m.

Notons que $\mathbf{|\varepsilon| = 0}$ et, pour toute lettre a, $\mathbf{|\varepsilon|_a = 0}$.

Exemples.

$\mathbf{A_1} = \{ 0, 1 \}$

$m_1 = 00101011$, $m_2 = 1101$

$|m_1| = 8$, $|m_2|_1 = 3$

Alphabet, mot sur un alphabet (3)

$\mathbf{A_2}$ = { a, b, c }

m_3 = baba, m_4 = bac

$|m_4| = 3$, $|m_3|_a = |m_3|_b = 2$

A_3 = { 0, …, 9, +, -, *, :, (,) }

m_5 = (12 + 4) * (71 – 14:5)

$|m_5| = 16$

A_4 = { si, alors, sinon, >, a, b, $\leftarrow$, +, 0, 1 , … }

m_6 = si a > b + 1 alors a $\leftarrow$ 0 sinon b $\leftarrow$ 10

$|m_6| = 15$

Concaténation de mots

Soient **u** et **v** deux mots de **A***. La **concaténation** de **u** et **v** est le mot, noté **u.v** ou plus simplement **uv**, obtenu en « collant » le mot **v** à la suite de **u**. Ainsi, $|uv| = |u| + |v|$ et, pour toute lettre **a**, $|uv|_a = |u|_a + |v|_a$.

On notera u^n le mot **u.u. … .u** (n fois), avec $u^0 = \varepsilon$.

Exemple. u = aba, v = ca, uv = abaca, u^2 = abaaba
$|u| = 3$, $|v| = 2$, $|uv| = 5$, $|u^2| = 6$

Pour tous mots **u**, **v** et **w**, nous avons :

- $\varepsilon u = u\varepsilon = u$ (ε est élément neutre)
- **u.vw = uv.w = uvw** (associativité)
- mais, en général, $uv \neq vu$ (non commutativité)

Préfixes, suffixes et facteurs

Soient **u** et **v** deux mots de **A***.

Le mot **u** est un **préfixe** du mot **v** s'il existe un mot **w** de **A*** tel que **v = uw**.

De façon similaire, le mot **w** est un **suffixe** du mot **v** s'il existe un mot u de **A*** tel que **v = uw**.

Le mot **u** est un **facteur** du mot **v** s'il existe deux mots $\mathbf{w_1}$ et $\mathbf{w_2}$ de **A*** tels que $\mathbf{v = w_1 u w_2}$.

Exemples.

- aba est un *préfixe* (et un *facteur*) de **abacabac** : **abacabac**
- bac est un *suffixe* (et un *facteur*) de **abacabac** : **abacabac**
- aca est un *facteur* de **abacabac** : **abacabac**

Quelques propriétés…

Proposition 1. Si **u**, **v** et **w** sont trois mots de **A***, alors

$$uw = vw \iff wu = wv \iff u = v$$

Lemme de Levi. Si **u** et **v** sont tous deux préfixes de **w**, alors **u** est préfixe de **v**, ou **v** est préfixe de **u** (les deux si **u = v**).

Théorème (de commutation). Si **u** et **v** commutent (c'est-à-dire sont tels que **uv = vu**), alors **u** et **v** sont deux *puissances* d'un même facteur :

$uv = vu \iff$ il existe un mot f de A* et deux entiers p et q , p,q $\geq$ 0,tels que $u = f^p$ et $v = f^q$.

Preuve du théorème de commutation…

- ($\Leftarrow$) Si $\mathbf{u = f^p}$ et $\mathbf{v = f^q}$, alors $\mathbf{uv = vu = f^{p+q}}$.
- ($\Rightarrow$) Par récurrence sur $\mathbf{N = |u| + |v|}$:
  - si $\mathbf{N = 0}$, alors $\mathbf{u = v} = \varepsilon$ et donc $\mathbf{u = v = u^0}$.
  - si $\mathbf{u} = \varepsilon$, alors $\mathbf{u = v^0}$ et $\mathbf{v = v^1}$ (similaire si $\mathbf{v} = \varepsilon$).
  - si $\mathbf{|u| = |v| > 0}$, alors $\mathbf{uv = vu}$ entraîne $\mathbf{u = v}$, par le Lemme de Levi, et donc $\mathbf{u = v = u^1}$.
  - sinon, comme **u** et **v** sont préfixes de $\mathbf{uv = vu}$, l'un est préfixe de l'autre (Lemme de Levi).

    Supposons, sans perte de généralité, que $\mathbf{v = uw}$.

    On a donc $\mathbf{uv = vu} \Rightarrow \mathbf{u(uw) = (uw)u}$, c'est-à-dire $\mathbf{uuw = uwu}$, et donc $\mathbf{uw = wu}$ (Proposition 1).

    L'hypothèse de récurrence permet alors de conclure.

Langages (1)

Un **langage** **L** sur un alphabet **A** est un sous-ensemble, fini ou infini, de **A***.

Par exemple, sur l'alphabet **A = {a, b}**, on peut définir les langages suivants :

- $\mathbf{L_1 = A^2 = \{ aa, ab, ba, bb \}}$,
- $\mathbf{L_2} = \{$ mots d'au plus quatre lettres ayant autant de **a** que de **b** $\}$

  $\mathbf{= \{\varepsilon, ab, ba, aabb, abab, abba, baab, baba, bbaa\}}$,
- $\mathbf{L_3} = \{$ mots ayant deux fois plus de **a** que de **b** $\}$ (ce langage est infini).

Langages (2)

Autres exemples :

- **A = { a, …, z }**

  L = { mots de la langue française **}**

- **A = {** mots de la langue française **}**

  L = { phrases correctes **}**

- **A = { … }**

  L = { programmes C++ syntaxiquement corrects **}**

- etc.

Opérations sur les langages (1)

Soit **A** un alphabet. On définit les opérations suivantes sur les langages définis sur **A*** :

Union.

$$L_1 \cup L_2 = \{ m \in A^* \mathbin{/} (m \in L_1) \text{ ou } (m \in L_2) \}$$

Exemple. *{ a, ba, ac } ∪ { b, c, ba, a } = { a, b, c, ba, ac }*

Intersection.

$$L_1 \cap L_2 = \{ m \in A^* \mathbin{/} (m \in L_1) \text{ et } (m \in L_2) \}$$

Exemple. *{ a, ba, ac } ∩ { b, c, ba, a } = { a, ba }*

Produit (de concaténation).

$$L_1.L_2 = L_1L_2 = \{m \in A^* / m = m_1m_2,\ m_1 \in L_1 \text{ et } m_2 \in L_2\}$$

Exemple. *{ a, b, ba} . { b, ab } = { ab, aab, bb, bab, baab }*

Opérations sur les langages (2)

Puissance.

$$L^0 = \{ \varepsilon \}$$

$$L^n = \{ m \in A^* \; / \; m = m_1 m_2 \ldots m_n,$$

$$\text{avec } m_i \in L \text{ pour tout } i,\ 1 \leq i \leq n \}, \text{ pour tout } n \geq 1$$

Exemple. $\{ a, ba \}^2 = \{ aa, aba, baa, baba \}$

Étoile et "plus".

$$L^* = L^0 \cup L^1 \cup \ldots \cup L^n \cup \ldots$$

$$L^+ = L^1 \cup L^2 \cup \ldots \cup L^n \cup \ldots$$

Exemples. $\{ a, bc \}^* = \{ \varepsilon, a, bc, abc, bca, aaa, aabc\ldots \}$

$\{ a, bc \}^+ = \{ a, bc, abc, bca, aaa, aabc\ldots \}$

$\Rightarrow$ Observons que $L^+ = L.L^*$